

# CAPACIDAD PREDICTIVA DE LAS RELACIONES ESTADISTICAS ENTRE LOS VOLUMENES DE PRECIPITACION DE LAS CUENCAS DE LA PENINSULA IBERICA Y LAS TEMPERATURAS DE LA SUPERFICIE DE LOS OCEANOS ATLANTICO Y PACIFICO DEL PERIODO 1946-1994

Elía Diez Muyo  
María López Bartolomé

Servicio de Variabilidad y Predicción del Clima. INM

## RESUMEN

El objetivo del trabajo es buscar relaciones entre anomalías atmosféricas y anomalías de las temperaturas superficiales del agua del mar (SST) con distintas estaciones de retardo. Para ello se utiliza el Análisis de Correlación Canónica (ACC) entre las anomalías de SST del Atlántico y del Pacífico y las anomalías de los volúmenes de precipitación de cada una de las cuencas de la Península Ibérica en el periodo 1946-1994. Las correlaciones significativas son utilizadas para predecir. La medida de la bondad de las predicciones, se obtiene calculando las curvas ROC y el valor económico para los terciles seco, normal y húmedo. Los resultados obtenidos indican que, existe una capacidad predictiva, respecto a las SST del Atlántico para la misma estación, en la cuencas de la vertiente atlántica en invierno y en el Segura en primavera. Con respecto a las SST del Pacífico, se muestra capacidad predictiva para la cuenca Sur con dos estaciones de retardo a principios de la primavera y simultanea para el Levante a finales del verano.

## 1. Introducción

La base científica en que se apoya la predicción estacional es que las anomalías de las condiciones de contorno de la atmósfera, que son, la menos en parte, predecibles para escalas de tiempo más largas que las de la predicción a corto plazo, pueden tener un impacto significativo sobre la variabilidad atmosférica. El efecto de las anomalías de las temperaturas de la superficie del mar (SST) es hacer algunos regímenes de tiempo más probables que otros, por ejemplo, cambiando la probabilidad estadística de ocurrencia de situaciones de bloqueo sobre una región dada de la atmósfera (Palmer y Anderson, 1994).

Utilizando métodos estadísticos, se pueden buscar en los registros históricos, las relaciones entre anomalías atmosféricas regionales y anomalías de SST para la misma estación y con una o más estaciones de retardo. Cuando se encuentran relaciones significativas las anomalías de SST pueden ser usadas para predecir anomalías atmosféricas. Tales predicciones no son deterministas, los métodos estadísticos proporcionan una estimación, basada en la experiencia pasada de las anomalías atmosféricas regionales que probablemente ocurrirán.

El Análisis de Correlación Canónica (ACC), técnica estadística lineal *multivariante* ampliamente utilizada (Barnett y Preisendorfer, 1987; Barnston y Smith, 1996), se utiliza para explorar las relaciones entre las anomalías de SST de los océanos Atlántico y Pacífico (los campos *predictores* son cuasi-globales) y las anomalías de los volúmenes de precipitación de cada una de las cuencas de la Península Ibérica en el periodo 1946-1994. Las correlaciones obtenidas son utilizadas para predecir. La medida de la validez que así se alcanza, se obtiene calculando las proporciones ROC y el valor económico para los terciles seco, normal y húmedo.

## 2. Datos

Los datos de las SST proceden del GISST2.2 proporcionado por el centro *Hadley*. Son valores mensuales, con un *grid* regular latitud-longitud de 1°, del periodo enero-1946 hasta diciembre-1994. Las zonas elegidas son las siguientes:

- Pacífico: 60°N-60°S y 140°E-70°W
- Atlántico: 60°N-50°S y 80°W-20°E

Los datos de los volúmenes de precipitación (VP) de las cuencas en las que se divide la Península Ibérica (Pirineo Oriental (0), Norte (1), Duero (2), Tajo (3), Guadiana (4), Guadalquivir (5), Sur (6), Segura (7), Levante (8) y Ebro (9)) proceden del I.N.M. Son valores mensuales para el periodo enero-1946 hasta diciembre-1994.

A partir de las series mensuales se calculan las series de anomalías estacionales tanto para las SST como para los VP. Las estaciones que se consideran son las siguientes:

- Invierno: diciembre-enero-febrero (DEF) y enero-febrero-marzo (EFM)
- Transición del invierno a la primavera: febrero-marzo-abril (FMA)
- Primavera: marzo-abril-mayo (MAM) y abril-mayo-junio (AMJ)
- Transición de la primavera al verano: mayo-junio-julio (MJJ)
- Verano: junio-julio-agosto (JJA) y julio-agosto-septiembre (JAS)
- Transición del verano al otoño: agosto-septiembre-octubre (ASO)
- Otoño: septiembre-octubre-noviembre (SON) y octubre-noviembre-diciembre (OND)
- Transición del otoño al invierno: noviembre-diciembre-enero (NDE)

### 3. Resultados del Análisis de Componentes Principales (ACP)

La finalidad del ACP es la de simplificar la estructura de los datos, para poder explicar, en pocas componentes, la mayor parte de la información que contienen las variables. El número de componentes principales que se retienen es el criterio establecido por Preisendorfer-Barnett.

#### 3.1 EOF de las SST del Pacífico

Se obtienen de 10 a 12 CP dependiendo de la estación que se considere. La configuración espacial de la 1ª CP explica entre el 20 y 39 % de la varianza total, presenta un máximo en el Pacífico ecuatorial y en la serie temporal se ven los años de ocurrencia del fenómeno ENSO (Figura 1).

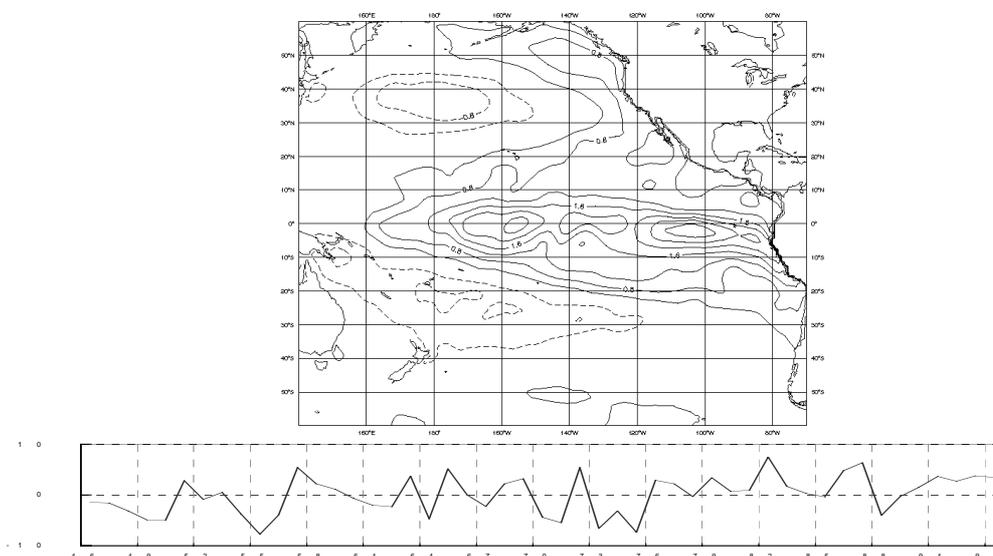


Figura 1: EOF1 del Pacífico y su componente principal para la estación de SON.

#### 3.2 EOF de las SST del Atlántico

Se obtienen de 12 a 13 CP dependiendo de la estación del año. La 1ª CP explica entre el 19 y 23 % de la varianza total con tres configuraciones espaciales diferentes. Desde el otoño (OND) hasta principios de la primavera (FMA) (Figura 2) se observa una configuración con un máximo en el sur del Golfo de Guinea y un mínimo al este de Terranova. En la primavera (MAM-AMJ-MJJ) y finales del verano-principios del otoño (ASO-SON) se observa una configuración similar a la anterior pero con un máximo en el sureste de Terranova y los valores negativos son más débiles y están desplazados hacia el NE del Atlántico Norte. En verano (JJA-JAS) se observan valores negativos en todo el Atlántico.

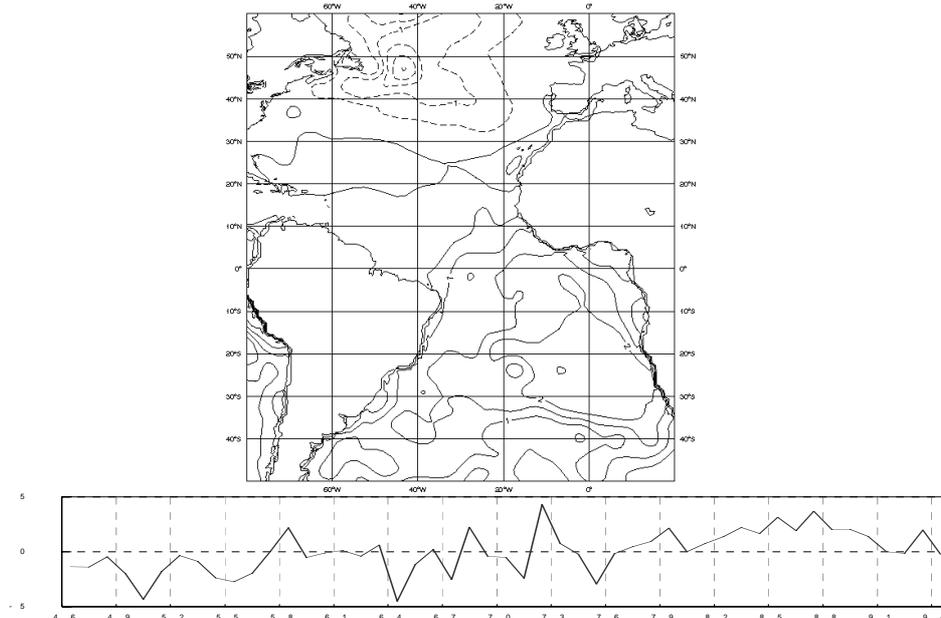


Figura 2: EOF1 del Atlántico y su componente principal para la estación de NDE.

#### 4. Resultados del Análisis de Correlación Canónica

El ACC es una técnica estadística *multivariante* que consiste en el estudio de las relaciones entre un conjunto de variables *predictores* (variables independientes,  $X_1, \dots, X_m$ ) y con otro conjunto de variables *predictando* (variables dependientes,  $Y_1, \dots, Y_n$ ). El objetivo principal es encontrar combinaciones lineales de  $m$  *predictores*  $U = u_1 X_1 + \dots + u_m X_m$  que tienen la máxima correlación ( $r$ ) con una combinación lineal de los  $n$  *predictandos*  $V = v_1 Y_1 + \dots + v_n Y_n$  siendo  $m \geq n$ .

Se ha realizado el ACC entre las componentes principales de las SST y los volúmenes de precipitación de cada una de las cuencas. Solo se consideran las correlaciones mayores de 60.

##### 4.1 Pacífico

Se encuentra una correlación simultánea en primavera y otoño. Una correlación con una estación de retardo del invierno a la primavera (NDE-DEF  $\rightarrow$  FMA-MAM) y de la primavera al verano (AMJ  $\rightarrow$  JAS). Las correlaciones más altas con dos estaciones de retardo son más bajas y aparecen aislados. Se encuentra una correlación con tres estaciones de retardo en el tercio norte de finales del invierno a finales del otoño.

##### 4.2 Atlántico

Se encuentra una correlación simultánea en la vertiente atlántica en todas las estaciones excepto en verano. Con la cuenca mediterránea las correlaciones más altas se dan en otoño. Con una correlación de una estación de retardo en las cuencas de la vertiente mediterránea del invierno a la primavera. Se encuentra una correlación con dos estaciones de retardo en las cuencas de la vertiente atlántica del otoño a la primavera. Se encuentra una correlación con tres estaciones de retardo en el tercio norte de la vertiente atlántica del invierno al otoño.

#### 5. Predicción de los volúmenes de precipitación a partir de las SST observadas

Como las variables canónicas  $U_i$ ,  $V_i$  tienen máxima correlación y suponiendo que ambas tienen varianza 1, la mejor predicción de  $V_i$  sobre  $U_i$  es

$$\hat{V}_i = r_i (U_i - \bar{U}_i) + \bar{V}_i \quad (1)$$

Se realizan predicciones en los casos en que las correlaciones son mayores de 65. Mostramos las predicciones con las SST del Atlántico simultaneas en DEF para el Tajo y Guadalquivir con una correlación de 65 (Figura 3) y para las SST del Pacifico simultanea en ASO en el Levante y con dos estaciones de retardo en FMA para la cuenca Sur ambas con una correlación de 71 (Figura 4).

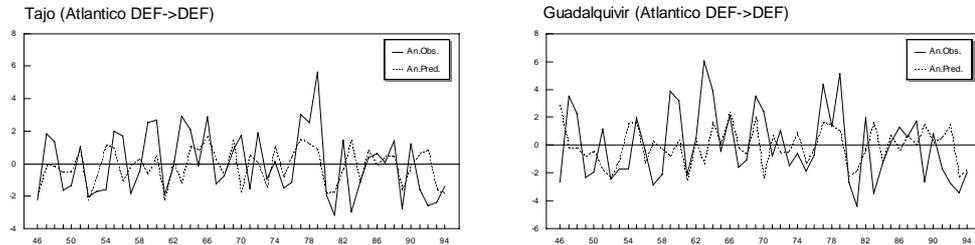


Figura 3: Predicciones con las SST del Atlántico simultaneas en DEF en las cuencas del Tajo y Guadalquivir.

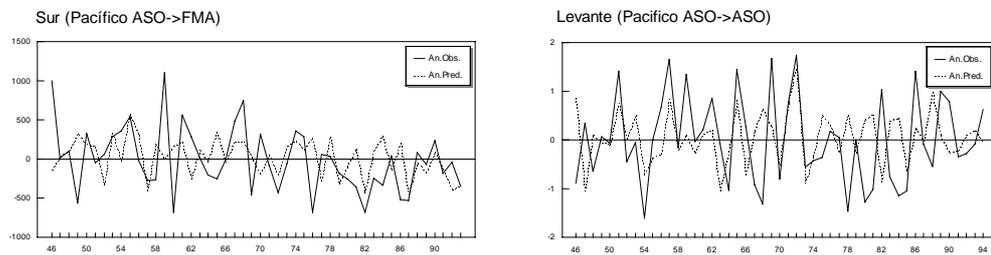


Figura 4: Predicciones con las SST del Pacifico en ASO para la cuenca Sur con dos estaciones de retardo y para el Levante simultaneas.

## 6. Validación del modelo de predicción (Técnica ROC)

Se utiliza el método de validación cruzada prediciendo cada año con el conjunto de datos después de haber extraído ese año. Para cada una de las estaciones consideradas se obtiene así un conjunto de predicciones de anomalías de volúmenes de precipitación de 49 años, que se valida respecto al mismo conjunto de 49 años de observaciones reales, dividiendo el conjunto de datos observados en terciles y utilizando la técnica ROC (Relative Operating Characteristic) para medir la fiabilidad de las predicciones para los tres casos: húmedo (h), normal (n) y seco (s).

La técnica ROC mide las proporciones de éxito y falsa alarma calculadas para una condición simple de predicción meteorológica, el evento E, que ocurre o no (se considera por tanto una predicción determinista de E). En nuestro caso, será que la anomalía predicha y observada caigan dentro del mismo tercil o no. Sobre una muestra suficientemente larga de predicciones independientes se puede construir la "matriz de contingencia" del modelo de predicción dando la frecuencia con que las anomalías observadas y predichas están dentro del tercil o no.

Predicción	Ocurrencia	
	NO	SI
NO	A	b
SI	C	d

Tabla 1: Matriz de contingencia

Basándose en estos valores se calcula la tasa de detección TD y la tasa de falsa alarma TFA de la siguiente manera:  $TD = \frac{d}{b + d}$   $TFA = \frac{c}{a + c}$  (2)

donde TD es la frecuencia con que el suceso ocurrió y fue predicho respecto a la frecuencia de ocurrencia del suceso y la TFA es la frecuencia con que el suceso fue predicho y no ocurrió respecto a la frecuencia de no ocurrencia del suceso. La fiabilidad de las predicciones se mide usando la escala de fiabilidad de Kuipers (KS). KS tiene la característica de que las predicciones aleatorias o constantes tienen asignado un

valor cero y las predicciones perfectas un valor 1. KS se define como  $KS = TD - TFA$ . Por consiguiente, el valor de la escala de Kuipers es igual al máximo valor que puede obtenerse con el sistema. Se han calculado las tasa de detección y falsa alarma y con ellas los valores KS.

## 7. Valor Económico de las predicciones

Aunque de las estadísticas ROC se puede obtener una medida objetiva de la fiabilidad de la predicción, es imposible saber que constituye un nivel de pericia útil para el usuario. Los resultados del análisis ROC relativas son la entrada a un modelo de decisión simple. Este modelo de decisión ha sido utilizado por Palmer et al. (1998) para definir el valor económico de las integraciones *ensemble* de *PROVOST*. El análisis del modelo de decisión es usado para definir una medida objetiva específica para cada usuario del valor económico de la predicción. Los resultados del análisis del modelo de decisión muestran que el valor económico de la predicción depende fuertemente de lo que cueste al usuario tomar precauciones en relación con la pérdida potencial si el episodio E ocurre. Considérese el caso del responsable de una decisión que solo puede tomar dos alternativas, emprender una acción o no hacer nada, y cuya elección depende exclusivamente de su creencia sobre si un episodio meteorológico dado E va a suceder o no. Emprender la acción implica un coste C independientemente de lo que suceda. Si el episodio ocurre realmente y no se ha emprendido acción alguna el interesado sufre una pérdida P. La "matriz de gasto" muestra las combinaciones posibles:

	Ocurrencia	
	NO	SI
Emprender acción	NO	0
	SI	C

Tabla 2: Matriz de gasto

El responsable de la decisión debe adoptar una estrategia que minimice a la larga sus pérdidas totales. Si únicamente dispone de información climatológica solo hay dos opciones, la primera será emprender siempre la acción protectora e incurrir siempre en un coste C y la segunda no emprender nunca la acción protectora y sufrir una pérdida P solo en el porcentaje  $\delta$  (ver matriz de contingencia) de ocasiones en que el episodio ocurre, por lo que la pérdida media es  $\delta P$ . Luego sin otra información más que la climatológica la estrategia óptima de actuación es emprender la acción si  $C < \delta P$  y no emprenderla nunca cuando no sea así. Pero el suministro de información complementaria en forma de predicciones puede permitir al responsable de la decisión reducir el gasto esperado. La proporción en que el gasto se reduce ofrece al usuario una medida del valor económico de las predicciones. Se define como valor económico V de un sistema de predicción a la proporción obtenida de reducción del gasto medio GM (gasto por unidad de pérdida) sobre la que se obtendría con una predicción perfecta. Así sería el valor máximo  $V=1$  para una predicción perfecta y el valor  $V=0$  para una predicción climatológica. Con  $V > 0$  el sistema es rentable. Utilizando los valores de la matriz de contingencia y la matriz de gasto se puede calcular el valor económico de la predicción en función de la razón C/P utilizando la siguiente fórmula:

$$V = \frac{\min(C/P, \delta) - TFA(1 - \delta)C/P + TD\delta(1 - C/P) - \delta}{\min(C/P, \delta) - \delta C/P} \quad (3)$$

siendo  $\delta$  la frecuencia climatológica de ocurrencia del suceso. A continuación se representan las gráficas del valor económico V frente a la relación C/P (coste/perdida).

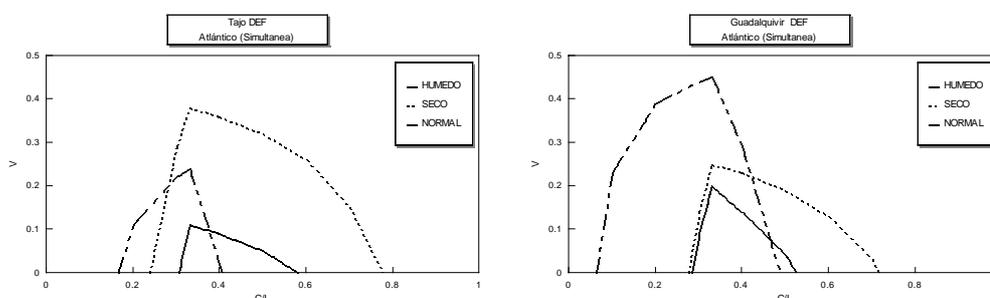


Figura 5: Valor económico para las predicciones con las SST del Atlántico simultaneas en DEF para el Tajo y el Guadalquivir.

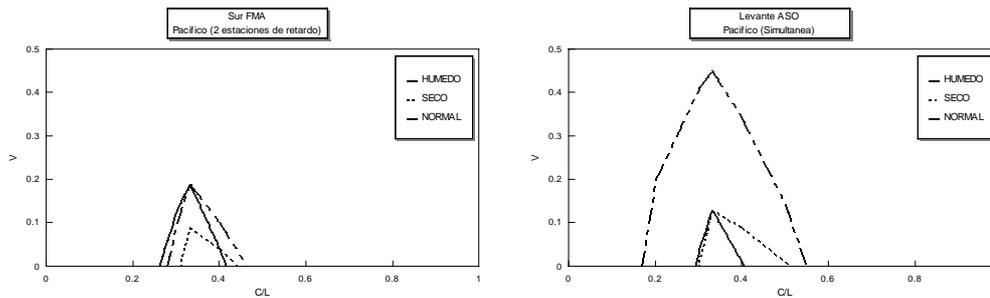


Figura 6: Valor económico para las predicciones con las SST del Pacifico en ASO para el Levante simultanea y el Sur con dos estaciones de retardo.

## 8. Conclusiones

Para las correlaciones mayores de 65 obtenidas del ACC se realizan predicciones, utilizando para cada año la ecuación obtenida con el resto de los años. La validez de las predicciones se calcula usando la escala de fiabilidad de Kuipers (se observan valores económicos positivos en los casos en que la KS es positiva) dando los siguientes valores positivos para las distintas cuencas, donde h, s y n son los terciles húmedo, seco y normal:

- Norte: Atlántico simultaneo en EFM (sn) y OND (hs).
- Duero: Atlántico simultaneo en DEF (s), EFM (hsn), SON (hsn) y OND (h).
- Tajo: Atlántico simultaneo en DEF (hsn), EFM (hsn), FMA (n), SON (s,n), OND (h) y NDE (h).
- Guadiana: el Atlántico simultaneo en DEF (hsn), EFM (hsn), OND (hsn) y NDE (hsn).
- Guadalquivir: Atlántico simultaneo en DEF (hsn), EFM (hsn), SON (hsn), OND (hsn) y NDE (hn).
- Sur: Pacífico con una estación de retardo en FMA (hs), MAM (h) y JAS(s), con dos estaciones de retardo en FMA (hsn). Con el Atlántico simultaneo en OND (n) y NDE (hsn), con una estación de retardo en MAM (hn) y con dos estaciones de retardo en AMJ (hs).
- Segura: Pacífico simultaneo en ASO (n), con el Atlántico simultaneo en MAM (hsn), NDE (hn) y con una estación de retardo en MAM (hsn).
- Levante: Pacífico en ASO (hsn), SON (sn), con dos estaciones de retardo en OND (s), con el Atlántico simultaneo en NDE (hn), y con una estación de retardo en OND (hs).
- Ebro: Atlántico simultaneo en OND (n) y NDE (h).

En resumen, existe una capacidad predictiva, respecto a las SST del Atlántico para la misma estación, en la cuenca del Duero en invierno (EFM) y otoño (SON), en la cuenca del Tajo en invierno (DEF, EFM), en la cuenca del Guadiana a finales del otoño y en invierno (OND, NDE, DEF, EFM), en la cuenca del Guadalquivir en invierno (DEF, EFM) y otoño (SON, OND), en la cuenca Sur a principios del invierno (NDE) y en la cuenca del Segura en primavera (MAM). Con respecto a las SST del Pacífico, para la cuenca Sur con dos estaciones de retardo a principios de la primavera (FMA) y simultanea para el Levante a finales del verano (ASO).

## BIBLIOGRAFIA

Barnett T.P. and Preisendorfer R., 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. Monthly Weather Review. Vol. 115. 1825-1849.

Barnston A.G. and Smith T.M., 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. J. of Climate. Vol. 9, 2660-2697.

Palmer T.N. et al., 1998: A Probability and Decision-Model Analysis of PROVOST Seasonal Multi- se observan valores económicos positivos en los casos en que la KS es positiva model Ensemble Integrations. Technical Memorandum No. 265. ECMWF.